Estimating Tremor in Vocal Fold Biomechanics for Neurological Disease Characterization

Pedro Gómez-Vilda, Víctor Nieto-Lluis, Victoria Rodellar-Biarge, Agustín Álvarez-Marquina, Luis M. Mazaira-Fernández, Rafael Martínez-Olalla, Cristina Muñoz-Mulas
Neuromorphic Speech Processing Lab
Centro de Tecnología Biomédica, Universidad Politécnica de Madrid
28223 Pozuelo de Alarcón, Madrid, Spain
e-mail: pedro@fi.upm.es

Mario Fernández-Fernández, Carlos Ramírez-Calvo
Ear, Neck and Throat (ENT) Services
Hospital del Henares
Avda. Marie Curie, s/n
28822 Coslada, Madrid, Spain

Abstract— **Neurological Diseases (ND) are affecting larger segments of aging population every year. Treatment is dependent on expensive accurate and frequent monitoring. It is well known that ND leave correlates in speech and phonation. The present work shows a method to detect alterations in vocal fold tension during phonation. These may appear either as hypertension or as cyclical tremor. Estimations of tremor may be produced by auto-regressive modeling of the vocal fold tension series in sustained phonation. The correlates obtained are a set of cyclicality coefficients, the frequency and the root mean square amplitude of the tremor. Statistical distributions of these correlates obtained from a set of male and female subjects are presented. Results from five study cases of female voice are also given.**

I. INTRODUCTION

Neurological Diseases (ND) are a major concern nowadays as life expectancy is growing in Western Countries. It is expected that a large percentage of elders will suffer from ND in their last years, demanding more resources for medical and social attention. For instance, *substantia nigra* cell decay responsible for Parkinson's Disease (PD) is about 5% per decade [1]. Many of these ND's affect voice and speech even at an early stage, when other symptoms are not yet evident [2],[3]. One of the most frequent perturbations found in voice in certain ND's as PD, is tremor. The aim of the present work is to develop a methodology to characterize tremor in ND's. Several approaches have been followed in this sense [4] with moderate results. Other approaches have successfully used classical acoustic perturbation correlates specifically for PD detection and grading [5]. The approach proposed here has been conceived to add semantic interpretation to tremor characterization in PD using phonation. It is based in removing articulation (vocal and nasal tract filters) on voiced segments of speech to estimate the phonation excitation (glottal source) [6]. In a second step, the power spectral density of the glottal source is matched to the transfer function of a second-order biomechanical model of the vocal folds [7]. In this way, estimates of biomechanical model parameters as the mass, stiffness and losses of the vocal fold inner body (*musculus vocalis*) and outer cover (*lamina propria*) are produced. The first hypothesis is that body and cover stiffness are induced by the stretching force exerted by laryngeal muscles on the vocal folds following neuromotor activity in the laryngeal nerve. Therefore vocal fold stiffness estimates would monitor neuromotor activity produced by midbrain centers responsible of vocal fold adduction and glottal closure. The second hypothesis is that perturbations of neuromotor activity producing tremor can be detected in time estimates of the vocal fold stiffness. The third hypothesis is that tremor-induced vocal fold stiffness perturbations can be characterized in frequency and amplitude by third-order auto-regressive systems. Section II is devoted to describe vocal fold body estimation and modeling by these auto-regressive systems. Section III describes the experimental framework to obtain statistical descriptions of tremor amplitude and frequency from a control set of Vocal Fold Organic Disease (OD) free subjects. Section IV shows the results for the control set as well as for several study cases including emotional and singing vibrato, spasmodic dysphonia and PD. Comparisons and differences are discussed. Section V highlights conclusions and describes future work.

II. CYCLICAL BEHAVIOR OF VOCAL FOLD STIFFNESS

The methodology to estimate tremor in vocal fold stiffness is summarized in Figure 1. The voiced speech trace $s_v(n)$ is inverse filtered to extract the glottal source $s_g(n)$ [6].

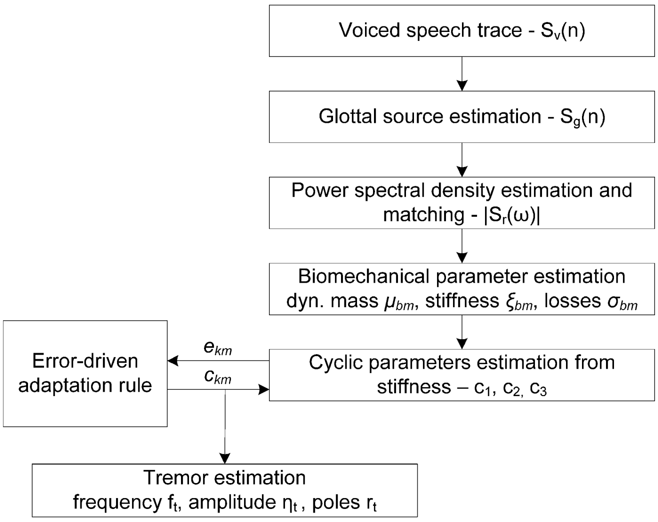

Figure 1 Estimating tremor correlates in vocal fold stiffness.

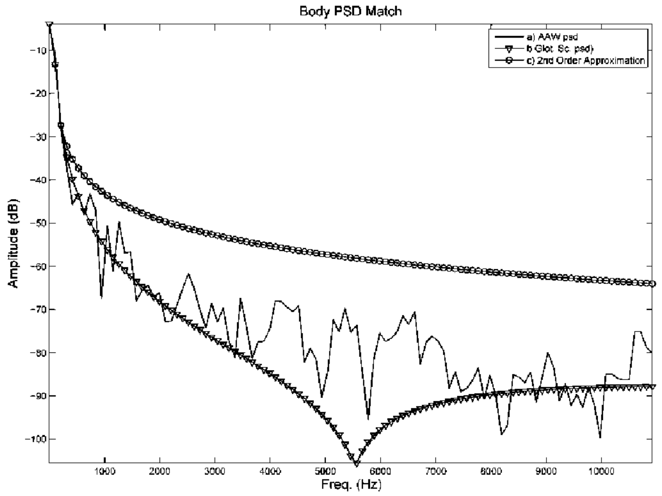

Figure 2 Matching the power spectral density of the glottal source against a second-order transfer function during a glottal cycle. Full blue line) Glottal source power spectral density. Red circles) Second-order approximation. Green triangles) Final approximation.

The glottal source power spectral density is estimated as $|S_r(\omega)|$. The envelope of $|S_r(\omega)|$ during a phonation cycle (interval between two successive closures of the glottis by the vocal folds) is matched against the transfer function of the second-order biomechanical model [6]. Defining its power spectral density as

$$\left\|S_r(\omega)\right\| = \left|\int_{-\pi}^{\pi} s_r(t) e^{-j\omega t} dt\right| \tag{1}$$

a cost function could be introduced to express the difference between the power spectral density and the transfer function of a biomechanical model transfer function [8], given as $T(\omega)$

$$L(\omega,\mu,\xi,\sigma) = \oint_{2\pi} \left(\left\|S_r(\omega)\right\| - \left\|T_c(\omega,\mu,\xi,\sigma)\right\|\right)^2 d\omega \tag{2}$$

where μ, σ and ξ stand for the estimates of each respective massive, viscous and elastic parameter of the body and cover biomechanics. Different matching functionals may be proposed for spectral fitting in (2). For instance assuming a single second-order functional as

$$\begin{aligned} T_c(\omega,\mu,\xi,\sigma) &= \left|Y_c\right|^2 = \left|\frac{V_c(\omega)}{F_c(\omega)}\right|^2 = \\ &= \left[\left(\omega\mu_c - \omega^{-1}\xi_c\right)^2 + \sigma_c^2\right] \end{aligned} \tag{3}$$

relating the cover mass velocity V_c with the applied force F_c in the frequency domain where Y_c is a mechanical trans-admittance, the process of optimization would imply the simultaneous fulfilling of the following conditions for the cover parameters

$$\frac{\partial L}{\partial \mu_c} = 0; \quad \frac{\partial L}{\partial \xi_c} = 0; \quad \frac{\partial L}{\partial \sigma_c} = 0; \tag{4}$$

Solutions for these conditions may be found either by forcing the derivatives of the functional L to zero deriving expressions for the three fitting parameters, or by adaptive gradient methods. The solution adopted in the present approach is based on fitting the glottal source power spectral density in Figure 2 (full line) by the transfer function (circles) given by (3) against the second-order approximation (triangles). As a result, a set of biomechanical parameters of the vocal fold body are produced for each phonation cycle m: μ_{bm} (dynamic mass), ξ_{bm} (stiffness) and σ_{bm} (losses). Later on, the stiffness parameter is modeled as an order-K autoregressive system

$$\xi_m = \sum_{i=1}^{K} a_i \xi_{m-i} + \varepsilon_m \tag{5}$$

where $\mathbf{a}=\{a_i\}$ are the model parameters and ε_m is the estimation error. This modeling is carried out by an adaptive lattice inverse filter [8]. Either the lattice filter pivoting coefficients $\mathbf{c}_{Km}$ or those of the equivalent transversal model $\mathbf{a}_{Km}$ may be used as cyclicality descriptors. Both sets of coefficients are related by the Levinson-Durbin iteration

$$\mathbf{a}_{km} = \mathbf{a}_{k-1m} - c_{km}\tilde{\mathbf{a}}_{k-1m} \tag{6}$$

$\tilde{\mathbf{a}}$ being vector $\mathbf{a}$ order-reverted. Pivoting coefficients are preferred for easy comparison, as they are pre-normalized to the interval (-1, 1). In the present case the three lowest-order pivoting coefficients $\{c_{1m}, c_{2m}, c_{3m}\}$ will be used as descriptors of the stiffness cyclicality pattern. In such case the relations among pivoting and model coefficients may be stated as

$$\begin{aligned} c_1 &= \frac{a_1 - a_2 a_3}{1 + a_2 - a_1 a_3 - a_3^2}; \\ c_2 &= \frac{a_2 - a_1 a_3}{1 - a_3^2}; \\ c_3 &= a_3 . \end{aligned} \tag{7}$$

Tremor may be described in terms of frequency, relevance and amplitude from the inverse model in the domain of z

$$\begin{aligned} H(z) &= \frac{1}{1 - \sum_{1}^{K} a_i z^{-i}} = \prod_{i=1}^{K} \frac{z}{z - z_i}; \\ z_i &= r_i e^{j\varphi_i}; \end{aligned} \tag{8}$$

where $\mathbf{z}=\{z_i\}$ are the poles of the transfer function $H(z)$, with modulus and phase given by r_i and φ_i. The phase of complex conjugate poles may be used to estimate tremor frequency as

$$f_{ti} = \frac{\varphi_i}{2\pi} f_0; \tag{9}$$

where f_0 is the phonation fundamental frequency. The robustness of the estimate will be given by the modulus of the pole (r_i). Another important parameter is the relative mean square tremor amplitude (rMSA), given by

$$\eta_t = \frac{\frac{1}{N_k} \sum_{n \in W_k} \left[\xi_{Kn} - \bar{\xi}_K\right]^2}{\bar{\xi}_K^2} \tag{10}$$

where N_k is the number of samples in the estimation window W_k. This model has important semantic properties. For instance, from (7) it may be shown that the closer the complex poles are to the unity circle ($r_i \rightarrow 1$), the closer will be c_1 to -1 ($c_1 \rightarrow -1$). Therefore c_1 could be used as a tremor mark. A question pertinent to the study is the robustness of c1 as a tremor mark, otherwise, to which extent complex pole estimates in (8) for K=3 given as

$$\begin{aligned} & z_2 = r_2 e^{j\varphi_2}; \quad z_3 = r_3 e^{j\varphi_3}; \\ & r_2 = r_3 \in \Re; \quad r_2, r_3 > 0; \\ & 0 \le \varphi_2 \le \pi; \quad \varphi_3 = -\varphi_2 \end{aligned} \tag{11}$$

are accurate enough. For such, let's consider the true ζ and estimated $\hat{\zeta}$ pole vectors and the estimation error ε_z

$$\begin{aligned} & \zeta = \{z_2, z_3\}; \quad \hat{\zeta} = \{\hat{z}_2, \hat{z}_3\}; \\ & \varepsilon_z = \hat{\zeta} - \zeta = \{\hat{z}_2 - z_2, \hat{z}_3 - z_3\}. \end{aligned} \tag{12}$$

The relative estimation error may be defined as the ratio between the absolute value of the estimation error and the modulus of the true pole vector

$$\begin{aligned} & \varepsilon_r = \frac{\|\varepsilon_z\|}{\|\zeta\|}; \\ & \|\varepsilon_z\| = \left[2\left(\hat{r}_2^2 + r_2^2 - 2\hat{r}_2 r_2 \cos(\hat{\varphi}_2 - \varphi_2)\right)\right]^{1/2}. \end{aligned} \tag{13}$$

This error has been estimated for the intervals ($0.5 < r_2 < 1$) and ($-\pi < \varphi_2 < \pi$) the results being plotted in Figure 3. It may be seen that the closer the pole to the unit circle, the smaller the error, independently of phase. This is a guarantee of robustness, as only relevant poles ($r_i \rightarrow 1$) will be taken into account.

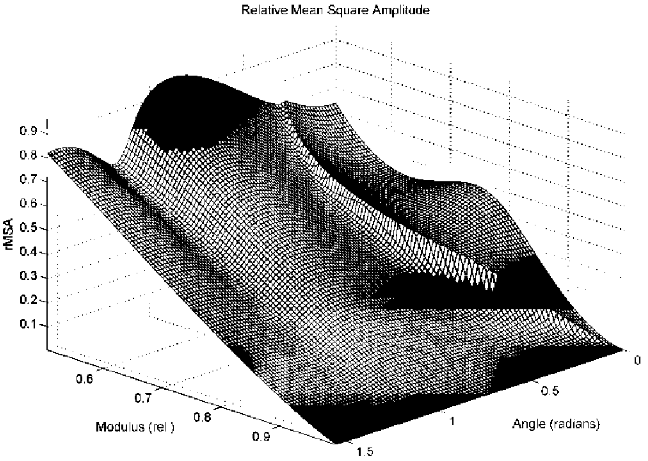

Figure 3 Pole estimation error depending on modulus and phase.

III. CONTROL SET STATISTICAL DISTRIBUTION

A database comprising 50 male and 50 female subjects has been used to evaluate general descriptive statistics for cyclicality parameters ($c_{1\text{-}3}$), tremor frequency (f_t), pole relevance (r_t) and rMSA (η_t). Other parameters of interest are the body and cover average stiffness (ξb, ξc), their unbalance between neighbor cycles ($\Delta\xi$b, $\Delta\xi$c), and their standard deviations ($\Sigma\xi$b, $\Sigma\xi$c). Each subject was first inspected by ENT (Ear, Neck and Throat) and Neurology Services to assess their normophonic condition (OD and ND free). Segments of /a/ 500 ms long were used in the study. The results of the statistical evaluations are given as histograms for each parameter separated by gender in Figure 4 and Figure 5. It may be seen that c_1 male and female distributions are skewed to -0.964 and to -0.978 for females (see for distribution statistical description in TABLE I, rel: relative units, N/m: Newton/meter).

TABLE I DISTRIBUTION STATISTICAL DESCRIPTION. Q: QUARTILES 1-3; (M/F): RESPECTIVE GENDER.

Qart./Par.	Q_1 (M)	Q_2 (M)	Q_3 (M)	Q_1 (F)	Q_2 (F)	Q_3 (F)
c_1 (rel)	-0.977	-0.962	-0.930	-0.984	-0.970	-0.934
c_2 (rel)	-0.016	0.303	0.463	-0.238	-0.007	0.149
c_3 (rel)	0.145	0.274	0.349	0.031	0.220	0.310
f_t (Hz)	2.716	3.819	4.902	4.492	6.002	8.909
r_t (rel)	0.898	0.935	0.953	0.884	0.914	0.947
η_t (rel)	0.006	0.010	0.013	0.007	0.010	0.016
ξ_b (N/m)	10.428	11.162	12.861	18.739	20.580	22.425
ξ_c (N/m)	6.856	9.651	13.922	20.490	29.582	35.043
$\Delta\xi_b$ (rel)	0.009	0.012	0.017	0.012	0.018	0.023
$\Delta\xi_c$ (rel)	0.028	0.040	0.060	0.037	0.050	0.072
$\Sigma\xi_b$ (N/m)	0.154	0.196	0.279	0.265	0.386	0.641
$\Sigma\xi_c$ (N/m)	0.330	0.564	1.248	1.355	1.891	3.171

Most cases show estimations of c_1 below -0.9, which means that the condition $r_i \rightarrow 1$ is fulfilled by a majority of cases (46/50 of males, 40/50 of females). This observation guarantees that these estimations are robust and reliable. The distributions of c_2 are more disperse, multimodal, and do not overlap completely between genders. The distributions of c_3 for both genders overlap better and are more compact. Confidence tests ensure equivalence of distributions in the case of c_1 and c_3 within a level of 0.05. This is not the case of c_2. Regarding tremor frequency f_t both distributions overlap well, although male distribution is centered toward slightly lower frequencies than female. In general most of the cases are within the range 2-10 Hz, with some cases above 10 Hz (5/50 of males, 12/50 of females). In any case 2 Hz is the lower limit granted by the segment duration for the analysis (500 ms). This situation prevents using this methodology in syllabic vowels, where duration seldom extends above 100 ms, but it may be used with fillers found in running speech, like the typical /uh/ and /ah, which may extend even to 1 second or more. The robustness of the estimates is granted by the proximity of the main pole to the unit circle as said ($r_t \rightarrow 1$). Most cases are well above 0.9 (47/50 in males, 41/50 in females). This grants that estimation errors are below 0.1 in absolute value accordingly to Figure 3. The tremor root mean square amplitude (rMSA or η_t) shows quite similar distributions in both genders, being between 0.5% and 2.5% in most cases (47/50 males, and 47/50 females). Cases out of these ranges could be considered anomalous. These distributions may be used to evaluate each analysis case when monitoring tremor in ND patients. As an example of use, five study cases are presented and discussed in the next section.

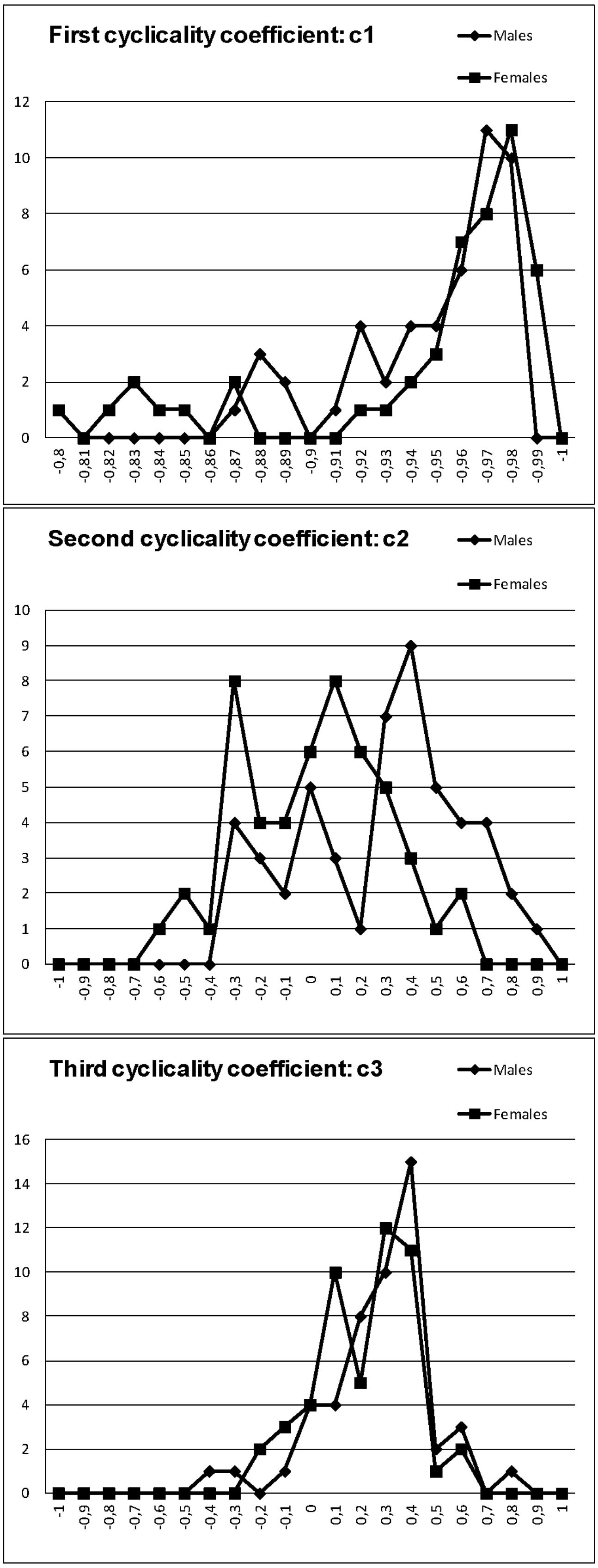

Figure 4 Histograms of the three cyclicity coefficients ($c_{1\text{-}3}$) for a population of normophonic male and female subjects.

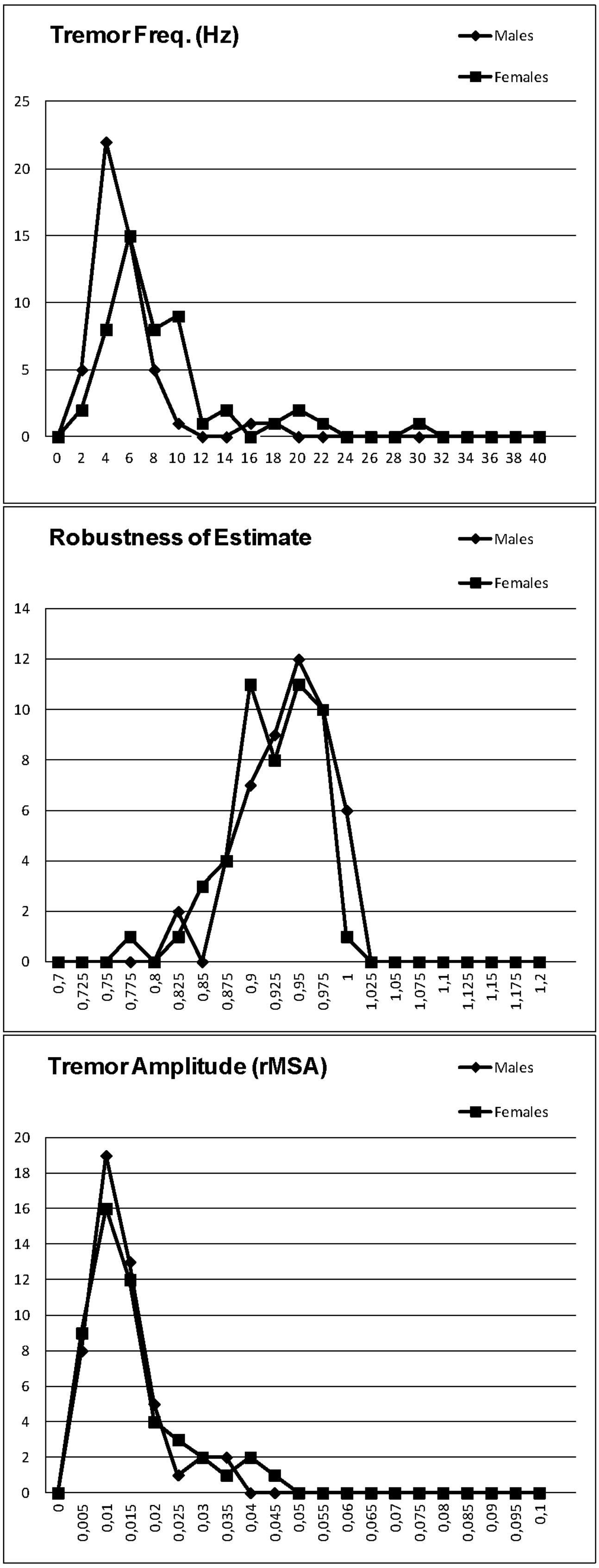

Figure 5 Histograms of tremor frequency, robustness and rMSA for a population of normophonic male and female subjects.

IV. RESULTS AND DISCUSSION

The study cases described are given in TABLE II. All subjects studied were female for a simple reason: female voice produces around twice glottal cycles per segment than male voice (around 100 cycles in a 500 ms segment), and therefore the estimates from a fixed duration segment are based in a larger sample size. The same study in male voice is left being carried out, and will not be shown here for the sake of brevity. The cases corresponded to a normophonic subject not showing tremor at first impression (100040, 28 years old), a normophonic subject showing emotional tremor (100350, 24 years old), a normophonic subject showing intentional tremor in singing (105001, 35 years old, singer, vibrato), a case of tremor in spasmodic dysphonia (100308, 45 years old), and a case of PD showing tremor (337523, 72 years old, stage 2).

TABLE II STUDY CASE DESCRIPTION

Case	**Gender**	**Condition**	**Description**
100040	F	Normal	No perceived tremor
100350	F	Normal	Emotional tremor
105001	F	Normal	Intentional vibrato
100308	F	Spasmodic	Low freq. tremor
337523	F	PD	High freq. tremor

The results for cyclicality and tremor estimates are given in TABLE III. The following meaningful observations may be pointed out:

- In all cases where tremor was significantly present (in bold) c_1 was below its first quartile (-0.984), whereas c2 moved to values well above its third quartile (0.149), and c3 did not present a clear tendency.
- It may be seen as well that case 100040 (supposedly tremor-free) showed tremor at 7.56 Hz but at a very low amplitude (ηt =0.5%). The cases with vibrato (105001) and PD (337523) showed moderate to large amplitude (2.1% and 3%, respectively, well above amplitude third quartile at 1.6%), and spasmodic dysphonia (100308) presented the largest amplitude (4.4%,).
- The robustness coefficient for these estimates was over its corresponding third quartile (0.947) except in the non-tremor normophonic case.

TABLE III STUDY CASES: CYCLICALITY AND TREMOR

Case	**c_1**	**c_2**	**c_3**	**f_t**	**r_t**	**η_t**
100040	-0.962	-0.184	0.054	7.566	0.902	0.005
100350	-0.977	-0.014	**0.343**	3.986	**0.959**	0.009
105001	**-0.988**	**0.579**	**0.472**	4.752	**0.957**	**0.021**
100308	**-0.993**	**0.603**	**0.255**	2.868	**0.986**	**0.044**
337523	**-0.986**	**0.517**	**0.501**	5.571	**0.979**	**0.030**

Stiffness estimates of the vocal fold body and cover give interesting results as well, as reported in TABLE IV (stiffness means and standard deviations ξ_b, ξ_c, Σ_b and Σ_c given in N/m, stiffness unbalances $\Delta\xi_b$ and $\Delta\xi_c$ given in relative units, see distribution quartiles in TABLE I). The following observations are relevant to the study:

- It may be seen that the body stiffness is in a range of Q1-Q3 for all cases except for PD (337523). Cover stiffness is also low in all cases except for the PD subject. This observation was confirmed in other studies including tremor and non-tremor in PD voice [11], pointing to the following hypothesis: tremor may be due to unstable adjustment of neuromotor activity between the correct tension to be applied to laryngeal muscles and the one really applied. The tremor would be a reaction of the neuromotor feedback in basal ganglia trying to re-adapt muscular tension to a more relaxed configuration, which is not always attained, and then tremor is a result of this failure. If feedback action is not successful laryngeal muscles may be over-tense and tremor will be lesser at the expense of a tenser phonation as revealed by the body and cover stiffness parameters (ξb, ξc).
- Body Stiffness unbalance ($\Delta\xi b$: difference between estimates from neighbor cycles) was below 1.8% except in cases of spasmodic dysphonia (100308) and PD (337523).
- Cover stiffness unbalance ($\Delta\xi c$) was also large in these two last cases, although it was extremely large in case 100040. This may be due to asymmetric vocal fold vibration rather than to tremor, this case being the one showing the lowest tremor amplitude, as commented.
- Finally the standard deviation of body and cover stiffness (Σb and Σc) show again larger values for spasmodic dysphonia and PD than in tremor-free, emotional and intentional vibrato.

TABLE IV VOCAL FOLD STIFFNESS AND DISPERSION

Case	**ξ_b**	**ξ_c**	**$\Delta\xi_b$**	**$\Delta\xi_c$**	**Σ_b**	**Σ_c**
100040	18.933	15.525	0.009	**0.068**	0.189	0.930
100350	20.948	20.748	0.010	0.022	0.335	0.955
105001	21.277	16.034	0.012	0.019	0.669	0.848
100308	20.271	16.638	**0.018**	**0.047**	**1.095**	**1.293**
337523	**25.186**	**29.323**	**0.023**	**0.054**	**1.044**	**3.403**

V. CONCLUSIONS

Through the present work a method to estimate vocal fold body and cover stiffness has been used which may be used in determining tremor in voice, which may be ultimately related to neuromotor activity of the laryngeal muscles. The estimation is based on the evaluation of a set of cyclicality coefficients related with third-order auto-regressive models of the stiffness time series. Pole estimates are used to compute tremor frequency and estimation robustness. Statistical distributions of the tremor frequency and amplitude as well as the first three cyclicality coefficients were obtained for a control set of normophonic male and female subjects. These distributions were used to evaluate five study cases of female voice including tremor-free, emotional, vibrato, spasmodic dysphonia and PD. It may be seen that the proposed methodology differentiates cases with low tremor (as in case 100040, where tremor-free was assumed, or in the emotional tremor case, 100350) from cases where strong tremor was present (as in the intentional vibrato case, 105001, spasmodic dysphonia, 100308, or PD, 337523). Another interesting conclusion is that vocal fold hypertension is also present in the PD case discussed. Besides, vocal fold stiffness statistical dispersion was larger in ND subjects than in normal ones, even if vibrato or tremor was present. This suggests a possible strategy to characterize ND phonation behavior, which would consist in fusing detection results using combinations from the following acoustic features: body and cover stiffness estimates, their unbalances and statistical dispersion, and tremor coefficients c_1 and c_2. An important study to be carried out regarding this strategy is the differential description of the population distributions of normophonic subjects, organic dysphonic and neurological disease dysphonic subjects in reference to the selected features. A study of feature relevance and redundancy in terms of mutual information contents may also help in improving detection performance [11]. The present study is to be extended to a large database of neurological disease cases for characterizing emotional, intentional or pathological cases in early detection and treatment monitoring [12].

ACKNOWLEDGMENT

This work has been funded by grants TEC2009-14123-C04-03 and TEC2012-38630-C04-04 from Plan Nacional de I+D+i, Ministry of Economic Affairs and Competitiveness of Spain. The authors wish to thank the collaboration of Drs. Bartolomé Scola and Daniel Poletti in the study of particular cases and in providing essential information on the databases collected at Hospital Universitario Gregorio Marañón of Madrid, Spain.

REFERENCES

[1] UK Dept. of Work and Pensions: http://www.dwp.gov.uk/publications/specialist-guides/medical-conditions/a-z-of-medical-conditions/parkinsons-disease. Accessed: January 10, 2013.

[2] T. Falk, W. Chan, and F. Shein, “Characterization of atypical vocal source excitation, temporal dynamics and prosody for objective measurement of dysarthric word intelligibility”, Speech Communication, vol. 54, no. 5,2012, pp. 622-631.

[3] Y. Yunusova, G. Weismer, J. Westbury, and M. Lindstrom, “Articulatory movements during vowels in speakers with dysarthria”, Journal of Speech, Language, and Hearing Research, vol. 51, 2008, pp. 596-611.

[4] Y. Pantazis, M. Koutsogiannaki and Y. Stylianou, “A novel method for the extraction of tremor”, Proc. of MAVEBA07 (Manfredi, C., Ed.), Florence University Press, 2007, pp. 107-110.

[5] A. Tsanas, “Accurate telemonitoring of Parkinson’s disease symptom severity using nonlinear speech signal processing and statistical machine learning”, D. Phil. Thesis, University of Oxford, UK, 2012.

[6] P. Gómez, R. Fernández, V. Rodellar, V. Nieto, A. Álvarez, L. M. Mazaira, et al., “Glottal Source Biometrical Signature for Voice Pathology Detection”, Speech Communication, vol. 51, 2009, pp. 759-781.

[7] M. P. deVries, H. K.. Schutte, A. E. P. Veldman and G. J. Verkerke, “Glottal flow through a two-mass model: Comparison of Navier-Stokes solutions with simplified models”, J. Acoust. Soc. Amer., Vol. 111, 2002, pp. 1847-1853.

[8] D. A. Berry, “Mechanisms of modal and non-modal phonation”, *J. Phonetics*, Vol. 29, 2001, pp. 431-450.

[9] J. R. Deller, J. G. Proakis, and J. H. L. Hansen, Discrete-Time Processing of Speech Signals, Macmillan Pub. Co., Englewood Cliffs, NJ, 1993.

[10] P. Gómez, V. Rodellar, V. Nieto, L. M. Mazaira, C. Muñoz, M. Fernández and E. Toribio, “Neurological Disease Detection and Monitoring from Voice Production”, LNCS, Vol. 7015, 2011, pp. 1-8.

[11] H. Peng, F. Long and C. Ding, “Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy”, IEEE Trans. On Pat. An. and Machine Int., Vol. 27, No. 8, 2005, pp. 1226-1238.

[12] J. Gamboa, F. J. Jiménez, A. Nieto, J. Montojo, M. Ortí, J. A. Molina, et al., “Acoustic Voice Analysis in Patients with Parkinson’s Disease Treated with Dopaminergic Drugs”, J. Voice, Vol. 11, 1997, pp. 314-320.